WOCCI 2017: 6th International Workshop on Child Computer Interaction
13 November 2017, Glasgow, Scotland, UK

Phontasia - A Phonics Game for German and its Effect on Orthographic Skills First Corpus Explorations

Kay Berkling

Cooperative State University, Karlsruhe, Germany
berkling@dhbw-karlsruhe.de

Abstract

This paper reports on first results in a study on orthography acquisition for children in German elementary school. One major aspect of German orthography concerns the marking of vowel duration, which occurs in every regular word in German. Therefore, mastery of this pattern has a large impact both on accurate spelling and very likely on reading ability. However, data shows that acquisition of this sub-skill appears difficult for kids to master, even beyond elementary school. In this paper, we explore a corpus of freely written text data that was collected on a weekly basis across several third grade classes in three different schools over a period of three months. Two of the schools participated in an intervention (iPad Phonics Game for German, called Phontasia). Studying the impact of the game on skill acquisition proved difficult because the ability to employ iPads in the classroom was correlated with socio-economic status of the schools. Results from VERA, a German standardized test, were obtained for several classes, allowing us to add one control group to the study. Preliminary results indicate that Phontasia intervention is worth pursuing in future studies to improve orthographic skills.

Index Terms: speech synthesis, human-computer interaction, serious games, orthography acquisition

1. Introduction

Acquisition of L1 German orthography in German schools is a topic that has not received much attention through quantitative analysis. Moreover, the situation is complicated by the number of school children in this country who do not speak German in their homes. It can be argued that schools are using L1 methods for L2 acquisition. However, effects of teaching material on acquisition are not evaluated in large-scale comparative studies. This explains why the IQB study [1] and VERA-3 results [2], published this year, resulted in an outcry of astonishment regarding the lack of performance in children's ability to spell, read and listen. VERA-3 is a standardized test for third grade that is taken by children all across the country to compare, among others, their skills in orthography.

VERA defines 5 levels of competency. The lowest two levels describe the ability to transcribe letters according to sound. At this level, many transcriptions are misspelled because German orthography is not flat, meaning, there are many-to-many correspondences between letters and phonemes. Different languages reflect various levels within the spectrum of orthographic depth, with English at the deep end, and languages like Finnish at the flat end. The German language, however, is deceptively located between the two and therefore poses some difficulties that need to be addressed during teaching in the first years [3]. Due to the current standard method of allowing students to write according to sound, this depth of orthography has been neglected during first grade [4, 5, 6] and might lead to bad habit formation. It is therefore not entirely surprising when the expected normal distribution across the five levels of difficulty has not been achieved. Instead of the expected 30%, the results report that 60% of children find themselves in the lower two levels of competency. Even more shocking is that 80% of children, whose daily language in their homes is not German, fall into the lower two levels [2].

The challenge of analyzing data for orthography acquisition is therefore compounded by the difficulty of obtaining sequenced teaching material that sets clear learning goals against which the writings can then be tested. Because the expected achievements of precisely described skills for various grade levels are not defined, research to measure their attainment is virtually impossible. The lack of research can be attributed both to the scarcity of openly available large corpora on children's writings and the hesitancy to use automation for evaluating massive amounts of such data at the required detail. Apart from the WISE tool [7], there is only one other semi-automated approach available for this work for the German language [8], neither of them in high demand.

To fill the gap of non-sequenced and unspecific skill definition, Phontasia was designed to teach reading and writing [9] in carefully crafted levels of increasing difficulty of orthographic principles. It follows the Phonics method that has been studied vastly in languages with a deeper orthography and may be well suited for a language like German that does not have a flat orthography [10].

This paper follows a sequence of publications with preliminary studies on smaller subgroups and has now been applied to a much larger group of children in third grade across several classrooms and schools. We will first briefly review some of the underlying principles to understand the work presented here in Section 2. Section 3 will introduce the game for to reader. Section 4 describes the collected corpus that will be explored in Section 5. Even though the study is much more extensive than our previous publications, results show how difficult this subject area is. We conclude with possible further steps in Section 6.

2. Background Knowledge

The goal of this section is to support a better understanding of the subsequent paper by offering an explanation of selected concepts.

2.1. Orthographic Depth

According to Ziegler et al. [11], the depth of the orthography [12] of a language is quantifiable through the entropy. This measure represents the difficulty of grapheme selection, given a phonological sequence (writing) and vice-versa, determining the correct phoneme given an orthographic sequence (reading).

The conditional entropy (H) for the process of reading $H(PHO|GRA)$ quantifies the indeterminacy of the phonetic

10.21437/WOCCI.2017-2

realization of a grapheme. PHO stands for the set of phonemes and GRA for the set of graphemes. The conditional entropy thus indicates how much information (measured in bits) is necessary in addition to the mere occurrence of a grapheme in order to realize the correct phoneme. The calculation is carried out according to Equation 1. In the case of a clear grapheme phoneme correspondence, $H(PHO|GRA)$ would be 0, so no additional information would be needed. The conditional entropy for the process of writing $H(GRA|PHO)$, is calculated by exchanging GRA and PHO. Here, the indeterminacy in the choice of the correct grapheme is quantified by means of the sound image.

$$H(PHO \mid GRA) = -\sum_{x \in GRA} p(x) * \sum_{y \in PHO} p(y \mid x) * log(p(y \mid x) \quad (1)$$

More detail on the consequences of this measure for German orthography acquisition and teaching materials can be read up in [6, 13]. The key result for our purpose in this paper is that orthography is not flat in German and writing is more difficult than reading.

2.2. Phonics

By specifying a word structure and the position of a phoneme within this structure, the choice of correct letters is facilitated for a learner by reducing the entropy. Therefore, word structures or patterns are a very important support factor in teaching reading and writing in languages with deep orthography. Phonics does exactly that, providing structure and small steps of increasing difficulties by growing the structure (and depth) slowly. Originally developed for English, phonics is a method that proposes explicit teaching of word patterns in sequences that minimize the phoneme-grapheme correspondence at the beginning and add minimal difficulty with each skill. In English, the first level is well known as the CVC Structure (“cat”, “hat”, “sat”, “mat”, or “set”, “get” etc.), which offers a fairly flat orthography within this pattern. Moving to the next level of difficulty, the “silent e” is added (“Kate”, “late”, “mate”, etc.). This pattern is no longer flat since letter <e> now has more than one phoneme, not taking word position into account. In this manner, each subsequent level builds on previously mastered skills.

2.3. Basic Levels in German Orthography

The most basic German word form, is the Trochee. It has two syllable, the first one stressed, the second one unstressed. The German language distinguishes between three major classes of Trochee that can be described as follows (C is a consonant phoneme, v,V are short and long vowel phonemes and R stands for reduction syllable rhyme, containing the letter “e” /ə/):

- 1 CVCR (b.e.t.en) or (b.ie.t.en)
- 2 CvCCR (b.e.s.t.en) or (B.i.l.d.er)
- 3 CvCR (B.e.tt.en) or (b.i.tt.en)

According to the entropy measurements of these word structures over a large set of German words and their pronunciation, we find that the level of difficulty increases from 1-3 in the above list. More about these levels can be read about in [6, 13]. Two of the most frequent spelling errors committed by children even beyond elementary school concern the incorrect spelling of the long “ie” in pattern 1 and the incorrect doubling of the consonant in pattern 3 that is used to mark the preceding vowel as short [13].

2.4. Phontasia - Phonics for German

Using levels of difficulty as described above and visualizing the placement of letters within restricted word structures, the process of reading and writing can be simplified because taking placement into account reduces orthographic depth. Phontasia practices these three patterns at a high frequency in a gameful manner, described in more detail in Section 3. Specifically, the above first three levels are trained with Phontasia. The beginning consonant yields only one consonant letter to choose from. In the first level, the player can choose only long vowels. Long and short vowels look the same with the exception of long /iː/ (“ie”) and short /i/ (“i”). Only one consonant letter fills the center of the trochee. The second level trains the short vowel, forcing the player to choose two consonant letters. Both levels restrict the players ability to choose words. They have to actively find words in their head that match the pattern. The general goal is to study whether this concept was grasped and transferred into written text at another occasion, unrelated to the game. If Phontasia intervention helps, then we expect a more significant decrease in the most frequent spelling errors in both test situation and freely written texts when compared to children who did not learn with Phontasia.

2.5. Mismatch between Teaching Material and Orthographic Depth

Currently, in German first grade materials, there are mainly 2-3 popular teaching methods in use. One such method is called “Lautiermethode” (sounding out method) and refers to the theory that words can be sounded out one letter at a time [14, 15]. By synthesizing the sounds, the child is expected to read any word. This works approximately for words like “Oma” or “Banane” (imported word) where one letter corresponds to one phoneme. Primary texts have an unnaturally high frequency of such words [5]. Unfortunately, this trains neural patterns that fail to generalize to typical German structures described above. Examples are given in Appendix D. Addressing reality of orthographic depth only in third grade presents a logical discontinuity with early teachings, the consequences of which remain unstudied.

2.6. Levels in VERA

VERA defines five skill levels that are broadly defined in Appendix D. The lowest level matches the described “Lautiermethode”. VERA provides results for expected performance in fourth grade, actual performance by country (Baden-Wrttemberg in our case) and performance per child and at classroom level. We have access to some of these results at the class level.

3. Phontasia - German Phonics

Phontasia is an iPad game that is designed to allow children to play while learning patterns of German orthography through grapheme-phoneme correspondence in context of increasingly complex word structures. The game allows the player to choose a letter/grapheme for each of the positions in a word. Once constructed, the word can then be read out using Apple native synthesis for any combination of letters. The interface is shown in Figure 1.

The game is designed to be addictive based on research in game design [16]. The key ingredients are immediate feedback, including fun death, a concept of failure that is consid-

ered hard fun. You can relate to the concept of "hard fun" by looking at the classic game of Ludo[1]. The player can lose three hearts while constructing 25 words correctly in order to open the next level of difficulty, allowing the player exactly two failures. When kids work in groups, you can witness heated debates at an orthographic meta-cognitive level about the correctness of the final word when they have one heart left (analogous to the last player in Ludo) and desperately need this last word to be correct in order to finish the level (analogous to reaching home in Ludo) or start from scratch. Each level is associated with a new set of vials that offers as a reward more difficult structures and therefore more word power as shown in Figure 2. Words that were not possible previously can now finally be constructed, thereby motivating the player. The reward in Level 3 is an even more complex system with a window tool to move up and down to choose between the two previous levels of word construction. Figure 3 depicts a higher level, where the player is given an additional vial for complex syllable onsets (containing more than one consonant).

Figure 1: *Level 1 of Phontasia practices simple 2-syllable words with long vowels.*

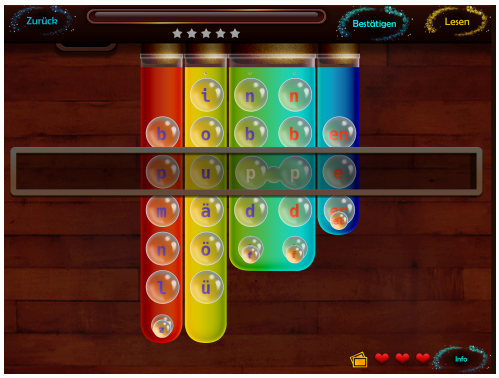


Figure 2: *Level 2 of Phontasia practices simple 2-syllable words with short vowel and single consonant phoneme.*

The game is personalized. Players who know the concept of a particular level can finish within minutes by correctly constructing 25 real words to open the next level. This can be achieved particularly quickly by using the blue vial, that contains morpheme endings (for example: "red.e", "red.en", "red.est" can be constructed by choosing only different endings, thereby speeding up the word construction and taking a deeper look at how conjugation and declination works.) Kids realize this feature after a while of playing. If, however, the letters are still new, the idea behind the alphabetic principle is still new, children can take many hours (spread over days/weeks) before they are able to finish all 25 words to open the next level. Therefore, the game caters to individual needs while the player is still having fun and getting exactly the practice they need at their own level of difficulty. Children that have a hard time managing the level can be observed to be very annoyed at putting down the game and impatient for their next chance. Their entire motivation is to unlock the next level, no matter how hard.

This game is a phonics game for German because it sequences letter-phoneme patterns from easy to more complex (increasing orthographic depth slowly), thereby offering practice that is increasingly difficult with explicit depiction of each level concept. It offers a multi-modal interface with sound, image and manual activity that are all tied into the teaching of a single new concept, catering to all learning modalities of the VAK model (visual, audio, and kinesthetic) [17].

Figure 3: *Level 6 of Phontasia practices minimal pairs for simple 2-syllable words, allowing the students to choose long or short vowels to create words.*

4. Corpus Description

The corpus collection in this study is described next. (More detail will be given in the corpus companion paper at LREC 2018 [18].)

4.1. Preliminary Study

In a previous study [19], a small group of 15 children were selected to work with us for one hour a week, using Phontasia and then writing a short text. At the end of the study, their performance was compared to their peers, who had not partaken in the intervention. While the results showed great promise, a larger study was needed, where Phontasia is used within the classroom and more children could participate. In this study, two schools and three classrooms were able to participate using Phontasia. A third school participated in writing texts but did not work with Phontasia.

4.2. Size of Corpus

The resulting corpus of children's writing has been collected in six different third grade classrooms during the 2016/2017

[1] ("Mensch rgere dich nicht") Getting your player thrown back to starting position by the opponent at the last second before reaching safety might make you mad, but does not stop you from playing. Essentially, this is the core component of what makes this game fun.

school year drawn from three schools (H2, ERK1, E2). It is important to note that we did not have the luxury to carefully select classes to design this study. These are the volunteers that we were lucky to have. Each class took part in the same pre- and posttest, consisting of the same dictation and list of pictures for which the children wrote the corresponding word.[2] During the course of nine weeks, children wrote texts for a provided picture prompt once a week (a similar setup for a previous corpus was already published [20]). In Appendix B, Figure 11 depicts an example prompt and Figure 12 shows a writing sample by a 9-year old boy from H2.

The average age of the children is 9 years old, ranging from 8-10 years of age. The data was collected from 3 different schools in the south of Germany. The corresponding regional dialect is visible in the written language ("B.e.d.e.g.e" /bədəgə/ for "B.e.tt.d.e.ck.e" /bɛtdɛkə/ - an effect of transcribing letters according to phonemes). 3 classrooms participated in the intervention with Phontasia (only E2 and ERK had access to iPads). The intervention consisted of children playing with Phontasia in the classroom once a week for 45 minutes. The writing prompt took place at a different time during the week.

Table 1: *Number of children and texts written by classroom.*

	H2			ERK		E2
	A	B	C	A	B	A
Texts	111	200	137,5	167	173	277
Kids	10	17	12	17	17	24

Table 1 gives the number of children participating in the study by classroom.[3] Table 2 lists the average number of words and types (unique words) found in children's texts for each of the weeks. It can be seen that there is a steady growth in text length observable for all classrooms. Observation by the transcribers state that the texts become more sophisticated, from lists of objects to descriptions and later on to full stories, including also more difficult vocabulary.

[2] The list of words for pre- and post-test are designed to elicit performance on a significant number of items with vowel length marking in regular words.

[3] 9 weeks + pre- + post-test + one additional text written several weeks after the post-test, which is not evaluated in this paper yields a total of maximally 12 texts to be collected for each kid.

Table 2: *Listing of number of words and types for each of the classrooms and each week, averaged across children. ER stands for ERK.*

Words	W1	W2	W3	W4	W5	W6	W7	W8	W9
H2.G3.KA	102	61	60	95	101	110	120	87	143
H2.G3.KB	88	95	81	111	96	124	138	115	96
H2.G3.KC	79	79	91	122	114	130	121	104	139
ER.G3.KA	48	85	42	99	87	106	105	74	103
ER.G3.KB	67	81	60	82	65	64	69	83	99
E2.KA	60	73	54	72	64	91	98	121	104

Types	W1	W2	W3	W4	W5	W6	W7	W8	W9
H2.G3.KA	56	34	35	61	61	63	67	57	77
H2.G3.KB	62	62	57	71	65	78	81	77	60
H2.G3.KC	51	46	61	72	69	77	71	64	82
ER.G3.KA	38	51	33	67	61	69	67	56	71
ER.G3.KB	45	52	43	53	48	45	47	59	68
E2.KA	45	50	42	51	49	62	63	80	73

Table 3: *% of children with complex language biography (indicating migration background) given the collected texts by classroom. (#Coll. = number of children whose texts were donated; #Class = estimated number of children in class)*

	%Migr.	#Migr.	#Coll.	#Class
H2.G3.KA	100%	10	10	26
H2.G3.KB	59%	10	17	22
H2.G3.KC	83%	10	12	25
E2.G3.KA	58%	14	24	26
ERK.G3.KA	53%	9	17	17
ERK.G3.KB	59%	10	17	17

4.3. Socio-Economic Background

According to the meta data, each of these classrooms has a percentage of children, whose language environment in their homes is not German. According to VERA, 80% of children who do not speak German at home perform at the lowest 2 levels in spelling. Therefore, we expect lower overall performance within classrooms with higher diversity of language biographies. The split of concerned children in this corpus is given in Table 3 as a percentage of obtained texts.

5. Data Exploration

In order to study orthography acquisition we look at the corpus. In addition, we have access to VERA test data that can be used to look at performance using a country-wide standardized test for comparison. However, it is well known (PISA) and recently supported again [21] that in Germany, school characteristics such as neighborhood and classroom intelligence have a significant impact on student performance. This makes comparative studies very difficult to impossible, not only across classrooms but even more so across both schools and classrooms.

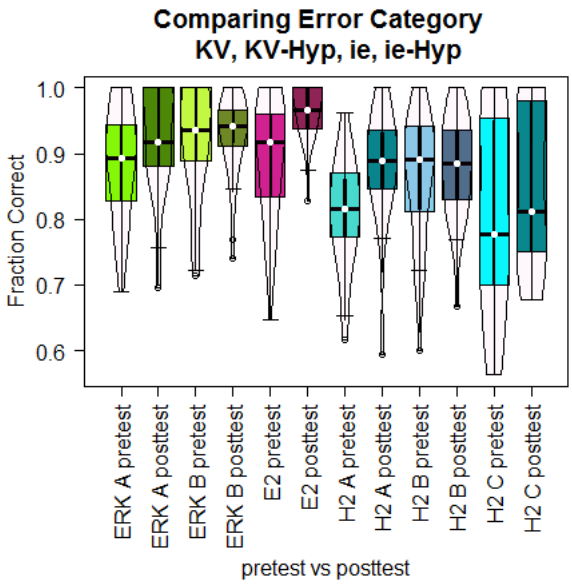


Figure 4: *Comparing fraction of correctly spelled error categories KV, ie and their over-generalization.*

Table 4: *Number of datasets (picture and dictation) for pre- and posttests (including both dictation and pictures.*

		Pretest	Posttest
ERK	KA	36	25
	KB	33	32
E2	KA	45	48
H2	KA	20	20
	KB	32	33
	KC	24	23

5.1. Spelling Error Categories

The collected corpus was analyzed with the automatic spelling error tagging tool WISE [7], giving us immediate access to a large number of spelling errors for any sized corpus. From this list, we select the long vowel "ie" (SIL_V_ie as in "bieten" /biːtən/) and the double consonants (SIL_V_KV as in "bitten" /bɪtən/) from pre- and post-test as well as for all freely written texts over the nine weeks to indicate skill level of correct marking of vowel length. In addition, these error categories can be summed up together with their respective over-generalization categories ($SIL\_V\_ie\_Hyp$, $SIL\_V\_KV\_Hyp$).

Looking at the number of correct writings divided by the total occurrence of the pattern in each category, we obtain a quotient value that is comparative across texts and children. This % correct measure should increase over the course of the school year, regardless of any intervention, simply due to the classroom effort. However, the hypothesis was that explicit teaching through the levels provided by Phontasia and a lot of personalized practice such as provided by the game, gives children enough skill improvement that there will be a noticeable effect for all children in both automaticity of implementing the orthographic concept in freely written text as well as the ability for the weak students to join the classroom level, an effect that had been shown in the previous study.

5.2. Pre- vs. Post-Test

The distribution changes are shown graphically in Figure 4 for pre- and post-test for the sum of the two selected vowel length marking skills, including over-generalizations. The depiction shows fraction of correctly written patterns for those categories and outlines their distribution (violin plot), drawing a bar for the median inside the box-plot. Note, that ERK1.G3.KA has 11 fewer data points in the post-test compared to the pre-test, see Figure 4, because a significant number of children left the class. All classrooms improve to different degrees in a test situation with isolated words, as quantified somewhat in Table 5 but the concept seems difficult to master in all classrooms. As expected, the H2 classrooms have much less fortunate statistics from the start. As a result, it is difficult to use H2 school as a control group for either E2 or ERK1 schools and this data does not provide a helpful view on the effect of Phontasia.

5.3. Weekly Writing

Next, we explore the data looking at the development of the two spelling error categories across the weeks of freely written texts. It seems that vacation may have an effect on spelling and length of the texts.[4] The progress for short vowel with double consonants is shown in Appendix A with Figures 7 and 8, comparing all six classrooms across the nine weeks of writing. Similarly, correct usage of "ie" is shown in Figures 9 and 10. While inspecting the graphs, it is important to keep in mind that the texts increased both in length and difficulty over the course of the nine weeks. While the quotient is normalized, it may still be more difficult to realize correct spellings while attention is being increasingly spent on story content. Looking at the change in spelling errors for both categories, the non-linear development of orthographic skills is more evident than a clear improvement in the skill. None of the classrooms for either error category show a clear slope of improvement. While the exam situation reflects that "ie" is correctly transcribed by E2 and ERK1 classrooms, this is not the case for freely written text. This view does not yield a clear picture of the effect of Phonatsia either.

[4]The week before vacation was as follows: H2 in week 5, ERK1 in week 3, and E2 in week 4.

Table 5: *Improvements in Figure 4 comparing means with p-value, as rough indicator, falsely assuming normal distribution. Three classrooms had significant improvements, two of those used Phontasia.*

		% improvement	p-value
ERK	KA	34%	0,08
	KB	18%	0,41
E2	KA	67%	0
H2	KA	36%	0,0198
	KB	17%	0,28
	KC	23%	0,26

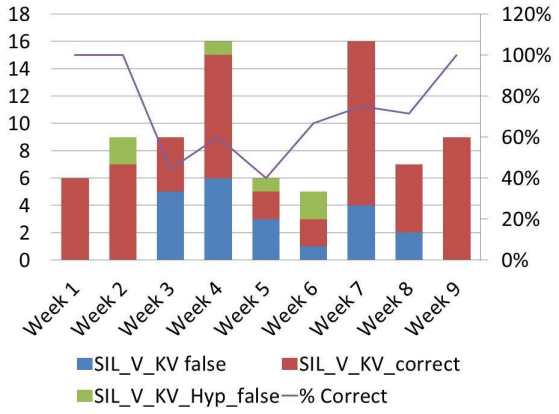


Figure 5: *Looking at orthographic development of single child.*

5.4. Child-based Observations

Figure 5 shows the progression for a random child in the corpus for SIL_V_KV. The graphic depicts the actual count for the number of times the pattern was spelled incorrectly (blue), correctly (red) and over-generalized (green). For this example, it can be seen that some over-generalization happens in the first half of the study. The number of errors increases as the texts become more complex but then decreases again towards the end of the study. Quantitative studies of child-relative improvements will need to take into account text complexity, spelling errors as well as their over-generalization patterns. Since every child's learning curve is individual, the impact of any interven-

tion seems impossible to relate with a single number, even less as an average across a classroom. However, the features depicted in Figure 5 contain patterns that can be recognized in variants across a number of children who used Phontasia. This level of child-dependent analysis will be necessary in order to demonstrate impact of Phontasia.

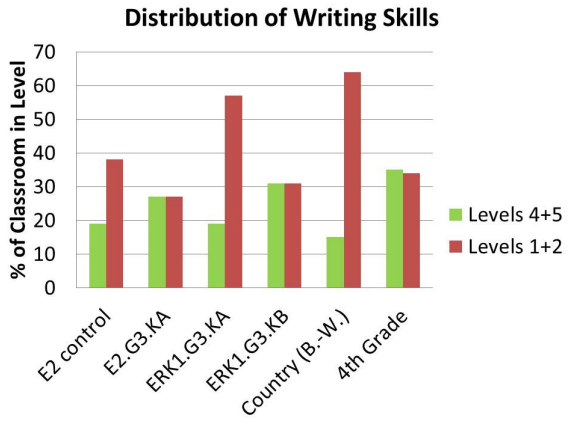


Figure 6: *Distributions across skill levels for various groups comparing % of classroom in lower Levels 1+2 and upper Levels 4+5 in VERA.*

5.5. VERA - Results

A more quantitative and standardized way of comparing classrooms that have used Phontasia to control groups may be by looking at the VERA results. In addition, VERA provides country-wide statistics for third grade as well as predictions for 4th grade performance as control groups. Note that while VERA reports that special education children were removed from the sample this could not be confirmed by the teachers. In addition, H2 results included second language learners and refugee children that have been in the country for less than a year (which is why the H2 VERA results are not usable here). Since H2 is not a valid control group, we look at another 3rd grade classroom in the E2 school that neither worked with Phontasia nor wrote our texts. Figure 6 shows the % distribution by combining the two highest and two lowest levels for each of the classrooms and compares these to fourth grade and country (Baden-Wrttemberg) performance[5]. It can be seen that ERK1.G3.KA is much closer (but slightly better) to the country average than all other classrooms. The other two Phontasia classes are more well-distributed and resemble the fourth grade results more closely than the E2 control class.

Quantitative comparisons are given in Table 6. We report the relative % increase of students in the top 2 levels (4,5) and the relative % decrease of students falling into the lower levels (1,2). The goal is to have a high increase in levels 4+5 and a high decrease for levels 1+2. Apart from ERK1.G3.KA, the other two Phontasia classes outperform the country average by far. When comparing the control group to the country results, we can see that E2 is better than the average school. Even so, two of the three Phontasia classes outperform this E2 control classroom.

[5] In comparison, Nordrhein-Westfalen, another state, has 55% (Levels 1+2) and 17.6% (Levels 4+5)

Table 6: *Comparing Improvements of classrooms. Large increase in % of children falling into high Levels 4+5 is good; large decrease in Levels 1+2 is good.*

	Cat. 4+5	Cat. 1+2
E2.G3.KA vs. country	80%	-58%
ERK1.G3.KA vs country	27%	-11%
ERK1.G3.KB vs. country	107%	-52%
control vs. country	27%	-41%
E2.G3.KA vs. E2 control	42%	-29%
ERK1.G3.KA vs. E2 control	0%	50%
ERK1.G3.KB vs. E2 control	63%	-18%

While these results are also not significant yet, given that there is only one control classroom in only one of the schools, they do indicate trends when compared to forth grade expectations and definitely encourage further study in this direction. However, as we can see by ERK1.G3.KA, not all classrooms benefit equally, or the change is not attributable to Phontasia. It is clear that getting a good control group when increasing the intervention groups is very difficult to accomplish across schools or even classrooms.

6. Future Work

We have looked at the long-term development of childrens' orthographic skills in third grade with and without intervention of a game that teaches German Phonics to children. We could show the environment-dependent and child-dependent non-linear path to acquisition that makes it very difficult to track improvement. The time-frame of nine weeks is probably also too short to draw full trajectories.

The text complexity, as it increases with each week, interacts with spelling errors and has to be studied in more detail. When looking at the VERA results, the intervention seemed effective but still anecdotal. There is still a lot of information in the data that we have not yet extracted or analyzed.

Finally, Phontasia is ideally suited for second half of first grade and the first half of second grade to obtain maximal impact. So far, we have not been able to find a classroom to work with us that early. Just as in sports and music, good habits are important from the start and wrong habits are hard to break once they have become automated. Even with a game like Phontasia that offers a lot of practice, it is disproportionately more difficult to unlearn bad habits in third grade than to teach correct ones early on.

This Corpus will be released through LDC in 2018 [22] and is described in an LREC 2018 paper ([18] expected to be published). The corpus adds to a previous comparative corpus (on the same text elicitations) that was also released through LDC [23].

7. Acknowledgements

Neither this work nor the corpus collection was deemed worth of funding. The University was able to sponsor some of the transcription costs. Thanks go to the kids, the parent who donate their work and the teachers who supported the study and the transcribers.

8. References

[1] P. Stanat, K. Böhme, S. Schipolowski, and N. Haag, "IQB-Bildungstrend 2015: Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten Ländervergleich," 2016.

[2] S. Blank and J. Schult, "Bericht VERA 3," 2017.

[3] K. Berkling and U. Reichel, "Der phonologische Zugang zur Schrift im Deutschen." in *Symposium Deutsch Didaktik, Sektion 7, Orthographie.*, Basel, CH, 2014.

[4] K. Berkling, U. Reichel, and R. Lavalley, "Untersuchung der Eignung von Fibeln für einen systematischen Schriftspracherwerb – Analyse von Fibeltexten durch automatische Sprach-verarbeitungsmethoden," in *Gesellschaft für Empirische Bildungs-forschung*, ser. GEBF, 2014.

[5] K. Berkling, "Item Presentation in Primers - An Analysis Based on Acquisition Research," in *Konferenz zur Verarbeitung Natürlicher Sprache*, ser. KONVENS, vol. 13, 2016.

[6] K. Berkling, R. Lavalley, and U. Reichel, "Systematic Acquisition of Reading and Writing: An Exploration of Structure in Didactic Elementary Texts for German," in *International Conference of the German Society for Computational Linguistics and Language Technology*, ser. GSCL. Gesellschaft für Sprachtechnologie and Computerlinguists, 2015, pp. 87–96.

[7] K. Berkling and R. Lavalley, "WISE – A Web-Interface for Spelling Error Recognition in German: A Description and Evaluation of the Underlying Algorithm," in *GSCL*, German Society for Computational Linguistics and Language Technology, Ed., 2015.

[8] R. Laarmann-Quante, K. Ortmann, A. Ehlert, M. Vogel, and S. Dipper, "Annotating orthographic target hypotheses in a german l1 learner corpus," in *The 12th Workshop on Innovative Use of NLP for Building Educational Applications, (BEA 2017), Copenhagen, Denmark, Sept 8, 2017*. EMNLP 2017 Workshops, 2017, pp. 444–456.

[9] K. Berkling and N. Pflaumer, "Phontasia - a Phonics Trainer for German Spelling in Primary Education," in *Workshop on Child Computer and Interaction WOCCI*, Singapore, 2014, pp. 33–38. [Online]. Available: www.wocci.org

[10] L. C. Ehri, S. R. Nunes, S. A. Stahl, and D. M. Willows, "Systematic phonics instruction helps students learn to read: Evidence from the national reading panels meta-analysis," *Review of educational research*, vol. 71, no. 3, pp. 393–447, 2001.

[11] J. C. Ziegler, D. Bertrand, D. Tóth, V. Csépe, A. Reis, L. Faísca, N. Saine, H. Lyytinen, A. Vaessen, and L. Blomert, "Orthographic depth and its impact on universal predictors of reading: A cross-language investigation," *Psychological science*, vol. 21, no. 4, pp. 551–559, 2010.

[12] P. H. Seymour, M. Aro, and J. M. Erskine, "Foundation literacy acquisition in european orthographies," *British Journal of psychology*, vol. 94, no. 2, pp. 143–174, 2003.

[13] K. Berkling and U. Reichel, "Wortstruktur, Orthographie und Didaktik:Die Relevanz der Vokallänge," in *Orthographische Kompetenz und Performanz im Spannungsfeld zwischen System, Norm und Empirie.*, ser. Thema Sprache: Wissenschaft für den Unterricht, System, Norm und Gebrauch - drei Seiten einer Medaille?, Ed. Hohengehren: Baltmannsweiler: Schneider, 2016, vol. 22, pp. 201–229.

[14] J. Reichen, *Hannah hat Kino im Kopf: Die Reichen-Methode Lesen durch Schreiben und ihre Hintergründe für LehrerInnen, Studierende und Eltern*, 5th ed. Hamburg: Heinevetter, 2008.

[15] H. Brügelmann, *Kinder auf dem Weg zur Schrift: Eine Fibel für Lehrer und Laien*, 9th ed., ser. Libelle: Wissenschaft. Bottighofen: Libelle-Verl., 2014.

[16] K. Berkling, H. Faller, and M. Piertzik, "Avoiding failure in modern game design with academic content - A recipe, an anti-pattern and applications thereof," in *CSEDU 2017 - Proceedings of the 9th International Conference on Computer Supported Education, Volume 2, Porto, Portugal, April 21-23, 2017.*, P. Escudeiro, G. Costagliola, S. Zvacek, J. O. Uhomoibhi, and B. M. McLaren, Eds. SciTePress, 2017, pp. 25–36. [Online]. Available: https://doi.org/10.5220/0006281800250036

[17] W. B. Barbe and M. N. Milone Jr, "What we know about modality strengths." *Educational Leadership*, vol. 38, no. 5, pp. 378–80, 1981.

[18] K. Berkling, "A 2nd corpus for children's writing with enhanced output for specific spelling patterns," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2018, Japan.*, 2018.

[19] K. Berkling, N. Pflaumer, and R. Lavalley, "German phonics game using speech synthesis - a longitudinal study about the effect on orthography skills Education, SLaTE 2015, Leipzig, Germany, September 4-5, 2015," in *Workshop on Speech and Language Technology in Education*, ser. SLaTE, vol. 6. ISCA(ISCA) International Speech Communication Association, 2015, pp. 167–172.

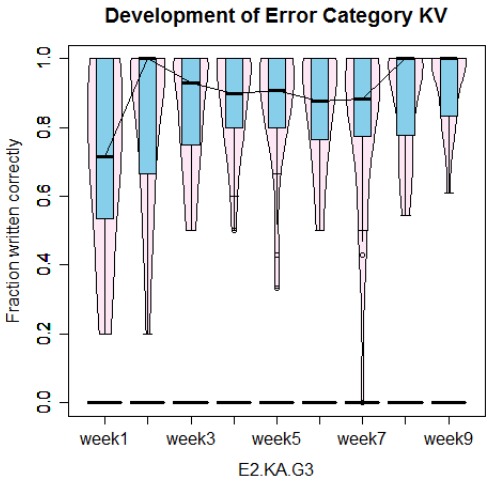
[20] K. Berkling, "Corpus for children's writing with enhanced output for specific spelling patterns (2nd and 3rd grade)," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016.*, N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds. European Language Resources Association (ELRA), 2016.

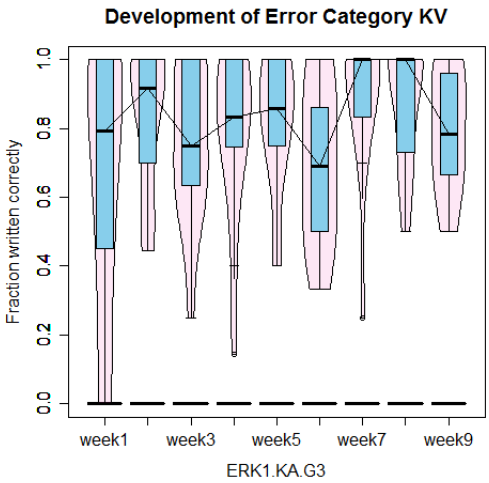
[21] A. von Suchodoletz, R. A. A. Larsen, C. Gunzenhauser, and A. Fsche, "Reading and spelling skills in german third graders: Examining the role of student and context characteristics," *British Journal of Educational Psychology*, vol. 85, no. 4, pp. 533–550, 2015. [Online]. Available: http://dx.doi.org/10.1111/bjep.12090

[22] Linguistic Data Consortium, "H2, E2, ERK1 Children's Text Corpus," https://catalog.ldc.upenn.edu/LDC2018T05, 2018, available: 2018-04.

[23] ——, "H1 Children's Text Corpus," https://catalog.ldc.upenn.edu/LDC2016T01, 2016, available: 2016-04.

A. Weekly Distributions





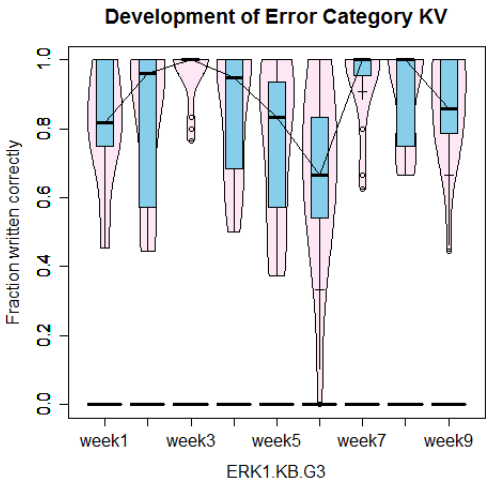
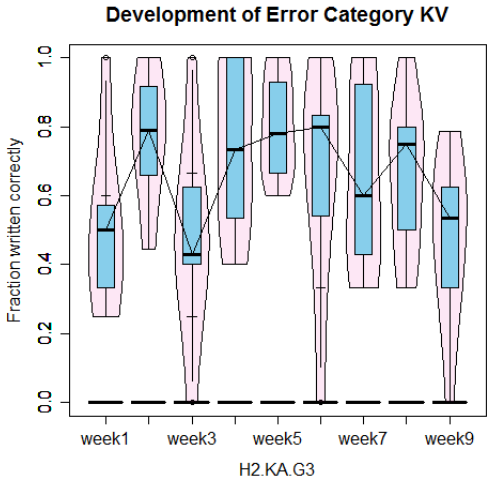
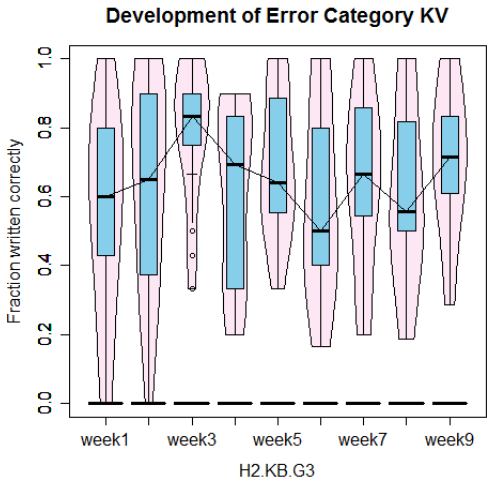


Figure 7: *Development of % correct short vowel marking through double consonant letters across weeks in classrooms with intervention.*





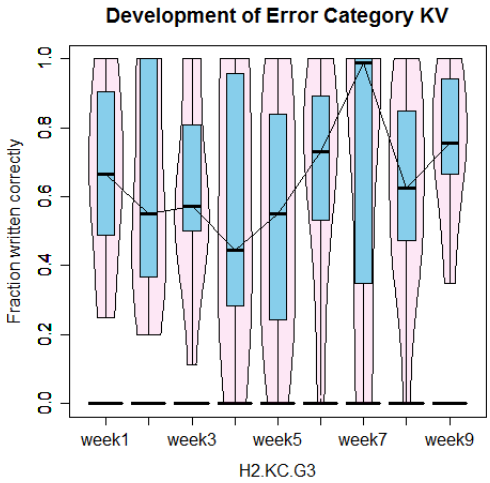
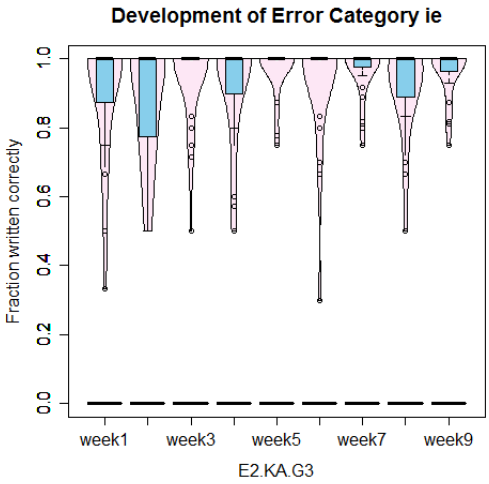
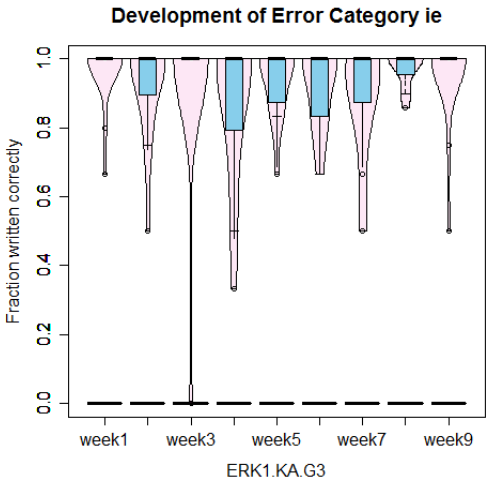


Figure 8: *Development of % correct short vowel marking through double consonant letters across weeks in classrooms without intervention.*





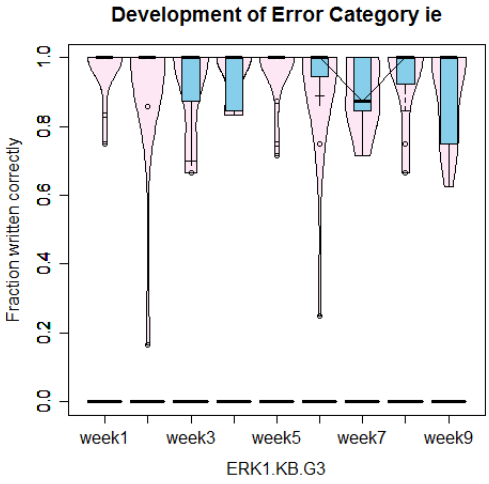
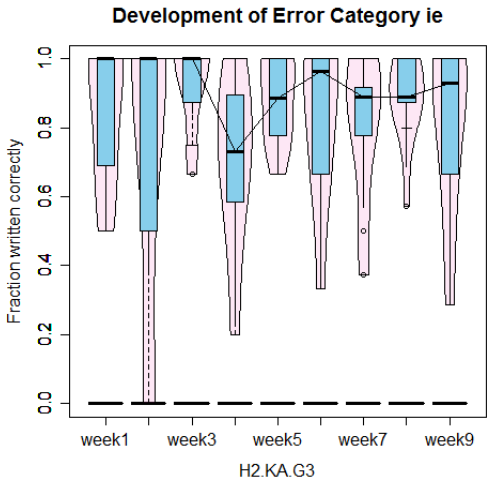
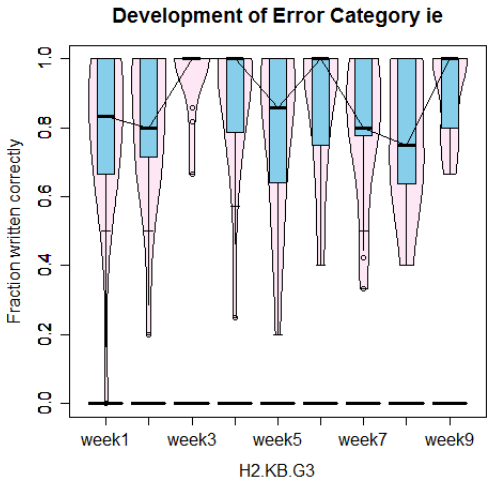


Figure 9: *Development of % correct written "ie" vowel across weeks in classrooms with intervention.*





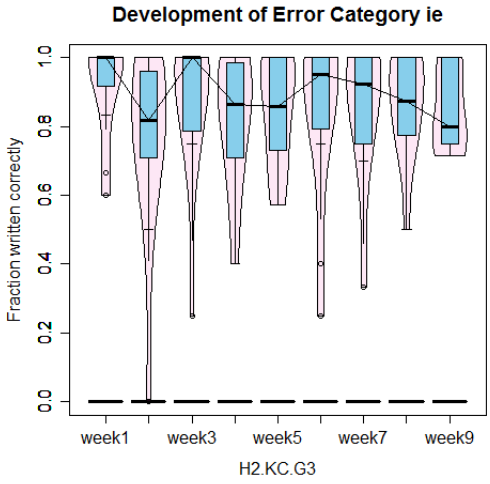


Figure 10: *Development of % correctly written "ie" vowel across weeks in classrooms without intervention.*

B. Data Elicitation

Figure 11: *Picture for text elicitation in the final week (week 9).*

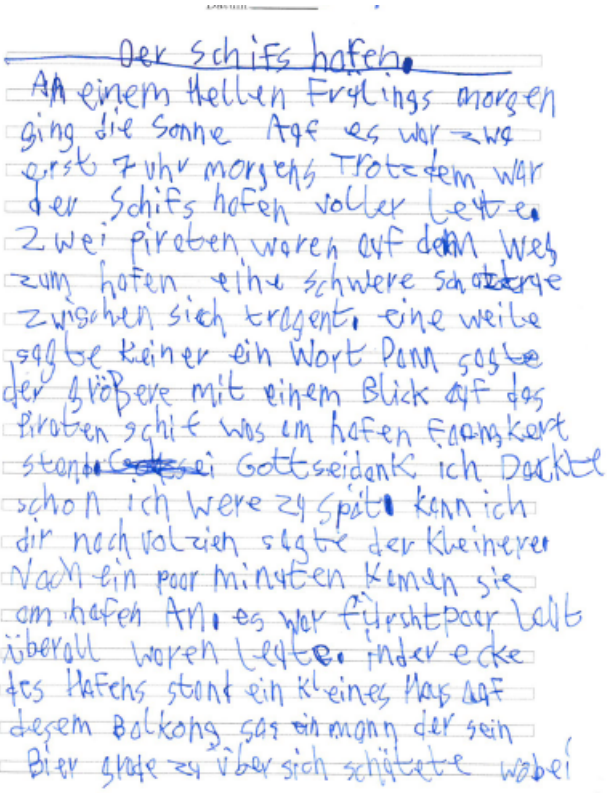


Figure 12: *Example text written by a 9-year old third grader whose mothertongue is German.*

Der Schiffshafen
An einem hellen Frhlingsmorgen ging die Sonne auf . Es war zwar erst 7 Uhr morgens , trotzdem war der Schiffshafen voller Leute . Zwei Piraten waren auf dem Weg zum Hafen , eine schwere Schatztruhe zwischen sich tragend . Eine Weile sagte keiner ein Wort . Dann sagte der grere mit einem Blick auf das Piratenschiff , was am Hafen verankert stand : Gott sei Dank, ich dachte schon , ich wre zu spt . Kann ich nachvollziehen, sagte der Kleinere . Nach ein paar Minuten kamen sieam Hafen an . Es war furchtbar laut . berall waren Leute . In der Ecke des Hafens stand ein kleines Haus , auf diesem Balkon sa ein Mann , der sein Bier grade zu ber sich schttete , wobei nur das Wenigste davon auch wirklich in seinen Mund drang . Ein Mann mit grantigem Unterkiefer und spitzer Nase angelte in der Ecke Fische und ein missbilligend guckender Pirat versuchte vergeblich seine Ziege ins Boot zu kriegen . Sie wollten grade die Trage ins Schiff hiefen , da rief jemand : He da ! Komm mal ! Die Beiden wirbelten herum . Gerufen hatte ein Mann , der tief in einer Karte steckte , die er offenbar las . Was ist , fragte der Grere und lie die Truhe fallen . Wir wissen , wo wir hin mssen , antwortete der Mann . Wohin? , fragte der Grere und eilte auf den Mann zu . In der Nhe von Costa Rica soll ein Schatz versteckt sein . Super!, sagte der Grere begeistert .

C. Example of Orthographic Depth

Using the so-called "Lautiermethode" (sounding out method), the child is expected to read any word by synthesizing the sounds of each letter. One such example, where this method falls short, is the letter <e> that can represent at least three semantically distinct pronunciations as a function of its position within syllable and word. Because the /ə/ in "Mutter" (/mʊtə/) sounds similar to the ending in "Oma" or "Lama" this results in spelling errors like "Muta", generalized from the unnaturally high frequency letter patterns presented to the child previously. Conversely, the written word "Mutter" is read as /muːt teːr/ and carries no semantic neural connection for the novice reader. In addition, German has graphemes that are made up of several letters and represent a corresponding phoneme jointly and in context dependent manner. For example, note the function of the letter "s/S" in the following three contexts for the same position in the words: <Sp> in "Spinne" (/ʃpɪnə/) vs. <Sch> in "Schnecke" (/ʃnɛkə/) vs. <s> "sagen" (/zaːgən/).

D. VERA Levels

In 2017, the writing skills are defined with respect to orthography. Level 5 spelling requires reasoning at several levels to obtain the correct grapheme sequence, for example by using syntax and morpheme knowledge. Level 4 requires knowledge of simple rules involving one of the above derivations ("Huser"). Level 3 uses simpler rules of orthography ("Hand"). Level 2 can use complex graphemes ("springen"), while Level 1 supposes a 1-1 correspondence between letters and phonemes ("Lama"). Level 1 skill reflects the view of German as a flat orthography, a 1-1 correspondence between phoneme and letter, and represents the lowest achievement level.